Geostatistics and Analysis of Spatial Data

Allan A. Nielsen

Abstract—This note deals with geostatistical measures for spatial correlation, namely the auto-covariance function and the semi-variogram, as well as deterministic and geostatistical methods for spatial interpolation, namely inverse distance weighting, radial basis functions (RBF) and kriging. Some semi-variogram models are mentioned, specifically the spherical, the exponential and the Gaussian models. Equations for RBF interpolation as well as simple and ordinary kriging (OK) are deduced. Other types of kriging are mentioned, and references to international literature, Internet addresses and state-of-the-art software in the field are given. A very simple example to illustrate the computations for OK and a more realistic example with height data from an area near Slagelse, Denmark, are given. A series of attractive characteristics of kriging are mentioned, and a simple sampling strategic consideration is given based on the dependence of the kriging variance of distance and direction to the nearest observations.

I. INTRODUCTION

OFTEN we need to be able to integrate point attribute information with vector and raster data which we may already have stored in a Geographical Information System (GIS). This can be done by linking the point information to a geographical coordinate in the data base. If we have lots of point data, a tempting alternative will be to generate an interpolated map so that from our point data we calculate raster data which can be analysed along with other sources of raster data.

This note deals with geostatistical methods for description of spatial correlation between point measurements as well as deterministic and geostatistical methods for spatial interpolation.

The basic idea in geostatistics consists of considering observed values of geochemical, geophysical or other natural variables as realisations of a stochastic process in the 2-D plane or in 3-D space. For each position \( r \) in a domain \( D \) which is a part of Euclidian space, a measurable quantity \( z(r) \) termed a stochastic function \( z(r) \) exists. \( z(r) \) is considered as a realisation of a stochastic variable \( Z(r) \). The set of stochastic variables \( \{ Z(r) | r \in D \} \) is termed a stochastic function. \( Z(r) \) has mean value or expectation value \( \mathbb{E}(Z(r)) = \mu(r) \) and auto-covariance function \( \text{Cov}(Z(r), Z(r + h)) = C(r, h) \), where \( h \) is termed the displacement vector. If \( \mu(r) \) is constant over \( D \), i.e., \( \mu(r) = \mu \), \( Z \) is said to be first order stationary. If also \( C(r, h) \) is constant over \( D \), i.e., \( C(r, h) = C(h) \), \( Z \) is said to be second order stationary.

This statistical view is inspired by work carried out by Georges Matheron in 1962-1963. It is described in for example [1], [2], [3] gives a good practical and data analytically oriented introduction to geostatistics. [4] is a chapter in a collection of articles which describe many different techniques and their application within the geosciences. [5] deals with geostatistics and other relevant subjects in the context of analysis of spatial data. Geostatistical expositions in a GIS context can be found in [6], [7]. [8] deals with multivariate geostatistics, i.e., studies of the joint spatial co-variation of more variables. The International Association for Mathematical Geology (IAMG) publishes i.a. the periodical Mathematical Geosciences (formerly Mathematical Geology) where many results on geostatistical research are published. State-of-the-art open source software may be found in GSLIB, [9], and Variowin, [10]. Other easily obtainable softwares are Geo-EAS and Geostatistical Toolbox. All these packages can be found at http://www-sst.unil.ch/research/variowin/ (or via a search engine). Also commercial geostatistical software exists.

This note which is inspired by [11] (see also [12]), in Section II deals with spatial correlation, specifically the auto-covariance function, the semi-variogram and some semi-variogram models are described. Section III deals with spatial interpolation including the deterministic methods inverse distance weighting and radial basis function interpolation along with a family of statistically based methods termed kriging. Here simple and ordinary kriging are dealt with in some detail. Section IV gives final remarks.

II. SPATIAL CORRELATION

This section mentions methods for description of similarity between measurements of natural variables in the 2-D plane or in 3-D space. Specifically the auto-covariance function and the semi-variogram are introduced. Also a relation between the two is given.

A. The Semi-Variogram

Consider two scalar quantities \( z(r) \) and \( z(r + h) \) measured at two points in the plane or in space \( r \) og \( r + h \) separated by the displacement vector \( h \). We consider \( z \) as a realisation of a stochastic variable \( Z \). The variability may be described b.m.o. the auto-covariance function (assuming or enforcing first order stationarity, i.e., the mean value is position independent)

\[
C(r, h) = \mathbb{E}[Z(r) - \mu][Z(r + h) - \mu] \]

The variogram, \( 2\gamma \), is defined as

\[
2\gamma(r, h) = \mathbb{E}[(Z(r) - Z(r + h))^2] \]

which is a measure for the expected squared difference between stochastic variables as a function of position and the displacement vector. In general the variogram will depend on the displacement vector \( h \) as well as on the position vector \( r \). The intrinsic hypothesis of geostatistics says that the semi-variogram, \( \gamma \), is independent of the position vector and that it depends only on the displacement vector, i.e.,

\[
\gamma(r, h) = \gamma(h) \]

If \( Z(r) \) is second order stationary (i.e., its auto-covariance function is position independent), the intrinsic hypothesis is valid whereas the opposite is not necessarily true.

If we assume or enforce second order stationarity the following relation between the auto-covariance function and the semi-variogram is valid

\[
\gamma(h) = C(0) - C(h). \tag{1}
\]

Note, that \( C(0) = \sigma^2 \), the variance of the stochastic variable.

Given a set of point measurements the semi-variogram may be calculated b.m.o. the following estimator, which calculates (half) the mean value of the squared differences between all pairs of measurements \( z(r_k) \) and \( z(r_k + h) \) separated by the displacement vector \( h \)

\[
\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} [z(r_k) - z(r_k + h)]^2.
\]
\(N(h)\) is the number of point pairs separated by \(h\). \(\hat{\gamma}\) is termed the experimental semi-variogram. Often we calculate mean values of \(\hat{\gamma}\) over intervals \(h \pm \Delta h\) for both length (magnitude) and direction (argument) of \(h\). Mean values for the magnitude of \(h\) \((h \pm \Delta h)\) are calculated in order to get a sufficiently high \(N(h)\) to obtain a low estimation variance for the semi-variogram value. Mean values over intervals of the argument of \(h\) are calculated to check for possible anisotropy. Anisotropy refers to the characteristic that the auto-covariance function and the semi-variogram do not behave similarly for all directions of the displacement vector between observations. This possible anisotropy may also be checked by calculating 2-D semi-variograms also known as variogram maps, [3], [13], [11], [10], [9].

### B. Semi-Variogram Models

In order to be able to define its characteristics we parameterise the semi-variogram b.m.o. different semi-variogram models. An often used model, \(\gamma^*(h)\), is the spherical model (here we assume or impose isotropy, i.e., the semi-variogram depends only on distance and not on direction between observations, and we denote by \(h\) the magnitude of \(h\))

\[
\gamma^*(h) = \begin{cases} 
0 & h = 0 \\
C_0 + C_1 \left[ \frac{3}{2} h - \frac{1}{2} h^3 \right] & 0 < h < R \\
C_0 + C_1 & h \geq R, 
\end{cases}
\]

where \(C_0\) is the so-called nugget effect and \(R\) is termed the range of influence or just the range; \(C_0/(C_0 + C_1)\) is the relative nugget effect and \(C_0 + C_1\) is termed the sill \((= \sigma^2)\). The parameters \(C_0\) and \(C_1\) are not to be confused with the auto-covariance function \(C(h)\). The nugget effect is a discontinuity in the semi-variogram for \(h = 0\), which is due to both measurement uncertainties and micro variability that cannot be revealed at the scale of sampling. The range of influence is the distance where covariance between samples ceases to exist; measurements taken further apart are uncorrelated.

Two other models often used are the exponential model (see Figure 4)

\[
\gamma^*(h) = \begin{cases} 
0 & h = 0 \\
C_0 + C_1 \left[ 1 - \exp \left( - \frac{3h}{R} \right) \right] & h > 0 
\end{cases}
\]

and the Gaussian model (see Figure 5)

\[
\gamma^*(h) = \begin{cases} 
0 & h = 0 \\
C_0 + C_1 \left[ 1 - \exp \left( - \frac{3h}{R} \right) \right] & h > 0. 
\end{cases}
\]

These latter two models never reach but approach the sill asymptotically. Due to its horizontal tangent for \(h \rightarrow 0\) the Gaussian model is good for describing very continuous phenomena.

Also other semi-variogram models such as linear and power functions are some times applied. To allow for so-called nested structures where the semi-variogram has different structures depending on the magnitude and possibly the direction of the displacement vector between observations, combinations of models may be useful.

The model parameters may be estimated b.m.o. iterative, non-linear least squares methods. These minimise the squared differences between the experimental semi-variogram and the model considered as a function of the vector of parameters \(\theta\), here \(\theta = [C_0 C_1 R]^T\)

\[
\min_{\theta} \| \hat{\gamma}(h) - \gamma^*(\theta, h) \|^2.
\]

For examples on an experimental semi-variogram and different models, see Figures 4 and 5.

Note, that \(C(0)\) is the auto-covariance function for displacement vector \(h = 0\), and that \(C_0\) is a parameter in the semi-variogram model.

### C. Examples

To illustrate the calculations Figure 1 shows a very simple example with three observations, \(z_1 = 1, z_2 = 3\) og \(z_3 = 2\) with (1-D) coordinates \(-2, -1\) og \(3\). The semi-variogram \(\hat{\gamma}\) with \(\Delta h = 1.5\) is calculated like this (lags are distance groups defined by \(h \pm \Delta h\))

<table>
<thead>
<tr>
<th>lag</th>
<th>(h)</th>
<th>(N)</th>
<th>(\hat{\gamma})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 &lt; h ≤ 3</td>
<td>1</td>
<td>1/2(1 - 3)^2 = 2</td>
</tr>
<tr>
<td>1</td>
<td>3 &lt; h ≤ 6</td>
<td>2</td>
<td>1/4((1 - 2)^2 + (3 - 2)^2) = 1/2</td>
</tr>
</tbody>
</table>

As another more realistic example Figure 2 shows a map with sample sites. Each circle is centered on a sample point and its radius is proportional to the quantity measured, in this case the height above the ground water in a 10 km × 10 km area near Slagelse, Denmark. Figure 3 shows a histogram for these data.

Figure 4 shows all possible pairwise squared differences as a function of distance between observations for the height data (assuming isotropy). Also an exponential variogram model estimated directly on this point cloud is shown. The nugget effect is 0 m², the effective range is 3,840 m and the sill is 840 m² (corresponding to 420 m² for the semi-variogram model).

Figure 5 shows the corresponding experimental semi-variogram. \(\Delta h\) is here 100 m and again we assume isotropy. Traditionally the first lag interval is half the size of the remaining lags, here 50 m. The experimental semi-variogram indicates

---

Fig. 1. Simple example with three observations.

Fig. 2. Sample sites, each circle is centered on a sample point, radius is proportional to the quantity measured, in this case the height above the ground water in a 10 km × 10 km area near Slagelse, Denmark.
that a Gaussian model may perform better than the exponential model in this case. Therefore a Gaussian model based on the experimental semi-variogram is shown also. The nugget effect is 18 m², the range is 1,890 m and the sill is 364 m².

III. SPATIAL INTERPOLATION

This section deals with deterministic types of interpolation such as inverse distance weighting and radial basis functions, and statistical types known under the joint name of kriging. Specifically equations for simple and ordinary kriging are deduced.

Other deterministic interpolation methods use (Delaunay) triangulation, Voronoi tessellation, regression analysis for determination of trend surfaces, minimum curvature etc., [15], [3].

A. Inverse Distance Weighting

Possibly the simplest conceivable way of carrying out interpolation consists of assigning the value of the nearest neighbour to a point where the value is unknown. An potential improvement consists of assigning higher weights to observations closer to the points to which we interpolate. An obvious way of doing this is to assign weights that are proportional to the inverse distance from the desired point to all \( N \) points entering into the interpolation. For the \( i \)th point we get the weight

\[
w_i = \frac{1/d_i}{\sum_{j=1}^{N} 1/d_j},
\]

where \( d_j \) is the distance from point \( j \) to the point to which we interpolate. This is readily extended to weighting with different powers, \( p > 0 \), of the inverse distance

\[
w_i = \frac{1/d_i^p}{\sum_{j=1}^{N} 1/d_j^p}.
\]

A.1 Examples

We now wish to interpolate to \( Z_0 \) at position \( r = 0 \) in Figure 1 b.m.o. inverse distance weighting, \( d_i \) is the distance from point \( Z_i \) to \( Z_0 \). We readily calculate the following weights

\[
\begin{array}{ccc}
 r & d_i & 1/d_i (1/d_i) / \sum(1/d_i) \\
-2 & 2 & 1/2 \quad 3/11 \quad (0.2727) \\
-1 & 1 & 1 \quad 6/11 \quad (0.5455) \\
3 & 3 & 1/3 \quad 2/11 \quad (0.1818) \\
\end{array}
\]

For different powers of \( d_i \) we get the weights

\[
\begin{array}{ccc}
 r & d_i & p = 0.1 \quad p = 2.0 \quad p = 10.0 \\
-2 & 2 & 0.3298 \quad 0.1837 \quad 0.0010 \\
-1 & 1 & 0.3535 \quad 0.7347 \quad 0.9990 \\
3 & 3 & 0.3167 \quad 0.0816 \quad 0.0000 \\
\end{array}
\]

We see that for low values of \( p \) the weights approach \( 1/N \) for all points used. For high values of \( p \) we get near a weight of one for the nearest neighbour.

B. Radial Basis Function Interpolation

Consider a linear estimate \( \hat{z}_0 = \hat{z}(r_0) \) at location \( r_0 \) based on \( N \) measurements \( z = [z_1, \ldots, z_N]^T = [z_1, \ldots, z_N]^T \) at locations \( [r_1, \ldots, r_N]^T \). Assume that each observation influences its surroundings in the same way in all directions and that the influence is expressed by some function \( \phi \) (the radial basis function, RBF) which depends on the distance \( h = ||r_0 - r_i|| \)
between locations \( r_0 \) and \( r_i \), only, \( \phi = \phi(||r_0 - r_i||) \). We shall look into choices of \( \phi \) shortly. Define the interpolated value

\[
\hat{z}_0 = \sum_{i=1}^{N} w_i \phi(||r_0 - r_i||),
\]

where \( w_i \) is the weight associated with location \( i \). Let us determine the \( w_i \) so that the interpolation becomes exact at the known locations \( r_j \), i.e.,

\[
z_j = \sum_{i=1}^{N} w_i \phi(||r_j - r_i||), \quad j = 1, \ldots, N.
\]

This makes up \( N \) equations with \( N \) unknowns, the \( w_i \), which can be written in matrix form

\[
\begin{bmatrix}
\phi(||r_1 - r_1||) & \cdots & \phi(||r_1 - r_N||) \\
\vdots & \ddots & \vdots \\
\phi(||r_N - r_1||) & \cdots & \phi(||r_N - r_N||)
\end{bmatrix}
\begin{bmatrix}
w_1 \\
\vdots \\
w_N
\end{bmatrix} =
\begin{bmatrix}
z_1 \\
\vdots \\
z_N
\end{bmatrix}
\]
or

\[
\Phi w = z.
\]

Often a polynomial \( P(r_j) = c_0 + c_1 x_j + c_2 y_j + \cdots = [1 \ x_j \ y_j \ \ldots] c \) where \( c = [c_0 \ c_1 \ c_2 \ \ldots]^T \) is the vector of coefficients for the polynomial is added to the interpolation at location (in two dimensions) \( r_j = [x_j \ y_j]^T \)

\[
z_j = P(r_j) + \sum_{i=1}^{N} w_i \phi(||r_j - r_i||),
\]

so that \( P^T w = 0 \). Here

\[
P =
\begin{bmatrix}
1 & x_1 & y_1 & x_1^2 & y_1^2 & x_1 y_1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_N & y_N & x_N^2 & y_N^2 & x_N y_N & \cdots
\end{bmatrix}
\]
defines the polynomial applied. Often a constant corresponding to the column of ones or a linear polynomial is used.

Solving the system of equations \( \Phi w = z \) under the constraint \( P^T w = 0 \) we get (see above on ordinary kriging and the Lagrange multiplier technique)

\[
\begin{bmatrix}
\Phi & P \\
P^T & 0
\end{bmatrix}
\begin{bmatrix}
w \\
c
\end{bmatrix} =
\begin{bmatrix}
z \\
0
\end{bmatrix}.
\]

The order of the polynomial may depend on the RBF chosen. References on RBF are [16], [17], [18].

B.1 Normalized RBF Interpolation

RBF interpolation comes in a normalized version also

\[
\hat{z}_0 = \frac{\sum_{i=1}^{N} w_i \phi(||r_0 - r_i||)}{\sum_{i=1}^{N} \phi(||r_0 - r_i||)}
\]

which leads to

\[
z_j = \sum_{i=1}^{N} \phi(||r_j - r_i||) = \sum_{i=1}^{N} w_i \phi(||r_j - r_i||), \quad j = 1, \ldots, N.
\]

for the weights. In matrix form this leads to the same matrix equation as above with the elements in the vector on the right hand side changed from \( z_j \) to \( z_j \sum_{i=1}^{N} \phi(||r_j - r_i||) \).

B.2 Choice of RBF

Often one of the following choices is made for \( \phi \)

- multiquadric: \( \phi(h) = (h^2 + h_0^2)^{1/2} \)
- inverse multiquadric: \( \phi(h) = (h^2 + h_0^2)^{-1/2} \)
- thin-plate spline: \( \phi(h) = h^2 \log(h/h_0) \) (which tends to 0 for \( h \) tending to 0), or
- Gaussian: \( \phi(h) = \exp(-\frac{1}{2}(h/h_0)^2) \),

where \( h_0 \) is a scale parameter to be chosen. Generally, \( h_0 \) should be chosen larger than a typical distance between samples and smaller than the size of the study area. The multiquadric is said to be less sensitive to the choice of \( h_0 \). Especially the Gaussian is sensitive to this choice.

With a Gaussian RBF you don’t need the polynomials mentioned above, with the thin-plate spline RBF a linear polynomial may be needed.

B.3 Shepard Interpolation

Shepard Interpolation

A special case for the normalized RBF interpolation consists of setting the matrix on the left hand side to a constant times the identity matrix. This corresponds to applying a \( \phi \) that tends to infinity for \( h \) tending to 0, and is finite for \( h > 0 \). This leads to setting the weights equal to the measurements \( w_i = z_i \), i.e., we needn’t solve the system of equations for \( w \). In this case \( \phi(h) = h^{-p}, 1 < p \leq 3 \) with appropriate handling for \( h = 0 \) is often used.

C. Kriging

Kriging (after the South African mining engineer and professor Danie Krige) is a name for a family of methods for minimum error variance estimation. Consider a linear (or rather affine) estimate \( \hat{z}_0 = \hat{z}(r_0) \) at location \( r_0 \) based on \( N \) measurements \( z = [z(r_1), \ldots, z(r_N)]^T = [z_1, \ldots, z_N]^T \)

\[
\hat{z}_0 = w_0 + \sum_{i=1}^{N} w_i z_i = w_0 + w^T z,
\]

where \( w_i \) are the weights applied to \( z_i \) and \( w_0 \) is a constant.

We consider \( z_i \) as realisations of stochastic variables \( Z_i, Z = [Z(r_1), \ldots, Z(r_N)]^T = [Z_1, \ldots, Z_N]^T \). We think of \( Z(r) \) as consisting of a mean value and a residual \( Z(r) = \mu(r) + \epsilon(r) \) with mean value zero and constant variance \( \sigma^2 \), \( \text{Var(}\epsilon)=0 \) and \( \text{Var(}\epsilon)=\sigma^2 \). For the linear estimator we get

\[
\hat{Z}_0 = w_0 + w^T Z.
\]

The estimation error \( z_0 - \hat{z}_0 \) is unknown. But for the expectation value of the estimation error we get

\[
E\{Z_0 - \hat{Z}_0 \} = E\{Z_0 - w_0 - w^T Z\}
\]

\[
= \mu_0 - w_0 - w^T \mu,
\]

where \( \mu_0 = \mu(r_0) \) is the expectation value of \( Z_0 \) and \( \mu \) is a vector of expectation values for \( Z \)

\[
\mu =
\begin{bmatrix}
\mu(r_1) \\
\vdots \\
\mu(r_N)
\end{bmatrix}.
\]

We want our estimator to be unbiased or central, i.e., we demand \( E\{Z_0 - \hat{Z}_0 \} = 0 \) or

\[
\mu_0 - w_0 - w^T \mu = 0.
\]
The variance of the estimation error is

\[
\sigma_E^2 = \text{Var}[Z_0 - \hat{Z}_0] \\
= \text{Var}[Z_0] + \text{Var}[w_0 + w^T Z] \\
- 2 \text{Cov}[Z_0, w_0 + w^T Z] \\
= \sigma^2 + w^T (C w - 2 \text{Cov}[Z_0, Z]),
\]

where \(C\) is the dispersion or variance-covariance matrix of the stochastic variables, \(Z\), entering into the estimation.

What is said in Section III-C so far is valid for all linear estimators. The idea in kriging now to find the linear estimator which minimises the estimation variance.

C.1 Simple Kriging

In simple kriging (SK) we assume that \(\mu(r)\) is known. From Equations 2 and 4 we get

\[ \hat{Z}_0 - \mu_0 = w^T (Z - \mu). \]

The weights \(w_i\) are found by minimising the estimation variance \(\sigma_E^2\). This is done by setting the partial derivatives to zero

\[ \frac{\partial \sigma_E^2}{\partial w} = 2 C w - 2 \text{Cov}[Z_0, Z] = 0, \]

which results in the SK system

\[ C w = \text{Cov}[Z_0, Z] \]

or

\[ \begin{bmatrix} C_{11} & \cdots & C_{1N} \\ \vdots & \ddots & \vdots \\ C_{N1} & \cdots & C_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \end{bmatrix}, \]

where \(C_{ij}, i, j = 1, \ldots, N\) is the covariance between points \(i\) and \(j\) among the \(N\) points, which enter into the estimation of point 0. \(C_{0j}, j = 1, \ldots, N\) is the covariance between point \(j\) and point 0, the point to which we interpolate. We get these covariances from the semi-variogram model (remembering Equation 1, \(\gamma(h) = C(0) - C(h)\) as the sill minus the value of the semi-variogram model for the relevant distance (and possibly direction) between observations. (Alternatively, the kriging system may be formulated b.m.o. the semi-variogram; to avoid zeros on the diagonal of \(C\) we prefer the covariance formulation for numerical reasons.) Here \(C_{ij}\) must not be confused with the semi-variogram parameters \(C_{0i}\) and \(C_{1i}\).

The minimised squared estimation error termed the simple kriging variance is

\[ \sigma_{SK}^2 = \sigma^2 + w^T (C w - 2 \text{Cov}[Z_0, Z]) = \sigma^2 - w^T \text{Cov}[Z_0, Z]. \]

In SK the mean value \(\mu(r)\) is known. In practice it is often assumed constant for the entire domain (or study area), or we must estimate it before the interpolation (or we must construct an interpolation algorithm which does not require knowledge of the mean field, see the next section).

C.2 Ordinary Kriging

In ordinary kriging (OK) we assume that the mean \(\mu(r)\) is constant and equal to \(\mu_0\) for \(Z_0\) and the \(N\) points that enter into the estimation of \(Z_0\). From Equations 3 and 4 we get

\[ \text{E}[Z_0 - \hat{Z}_0] = \mu_0 (1 - w^T 1) - w_0 = 0 \]

for any \(\mu_0\). \(1\) is a vector of ones. This is possible only if \(w_0 = 0\) and \(w^T 1 = 1\).

The weights \(w_i\) are found by minimising \(\sigma_E^2\) under the constraint \(w^T 1 = 1\). A standard technique for minimisation under a constraint is introducing a function \(F\) with a so-called Lagrange multiplier (here \(-2\lambda\)) which we multiply by the constraint set to zero and then minimising

\[ F = \sigma_E^2 + 2 \lambda (w^T 1 - 1) \]

without constraints. Again the partial derivatives are set to zero

\[ \frac{\partial F}{\partial w} = 2 C w - 2 \text{Cov}[Z_0, Z] + 2 \lambda = 0, \]

which results in the OK system

\[ C w + \lambda 1 = \text{Cov}[Z_0, Z] \]

or

\[ \left[ \begin{array}{ccc} C_{11} & \cdots & C_{1N} \\ \vdots & \ddots & \vdots \\ C_{N1} & \cdots & C_{NN} \end{array} \right] \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \end{bmatrix}. \]

The values requested for \(C_{ij}\) are found as described in the previous section on SK.

The minimised squared estimation error termed the ordinary kriging variance is

\[ \sigma_{OK}^2 = \sigma^2 + w^T (C w - 2 \text{Cov}[Z_0, Z]) = \sigma^2 - w^T \text{Cov}[Z_0, Z] - \lambda. \]

OK implies an implicit re-estimation of \(\mu_0\) for each new constellation of points. This is an attractive property making OK well suited for interpolation in situations where the mean is not constant (i.e., in the absence of first order stationarity).

C.3 Examples

Let us consider the data in Figure 1 again. We now wish to interpolate to the position \(r = 0\) b.m.o. ordinary kriging. To carry out the calculations we use a stipulated semi-variogram based on the spherical model with \(C_0 = 0\), \(C_1 = 1\) and \(R = 6\). Remembering Equation 1, \(C(h) = C(0) - \gamma(h)\), this gives the auto-covariance function (in this case where \(C_0 + C_1 = 1\) this is the same as the auto-correlation function)

<table>
<thead>
<tr>
<th>(h)</th>
<th>(\gamma(h))</th>
<th>(C(h))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>1</td>
<td>0.2477</td>
<td>0.7523</td>
</tr>
<tr>
<td>2</td>
<td>0.4815</td>
<td>0.5185</td>
</tr>
<tr>
<td>3</td>
<td>0.6875</td>
<td>0.3125</td>
</tr>
<tr>
<td>4</td>
<td>0.8519</td>
<td>0.1481</td>
</tr>
<tr>
<td>5</td>
<td>0.9606</td>
<td>0.0394</td>
</tr>
<tr>
<td>6</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Therefore the OK system looks like this

\[ \begin{bmatrix} 1.0000 & 0.7523 & 0.0394 & 1 \\ 0.7523 & 1.0000 & 0.1481 & 1 \\ 0.0394 & 0.1481 & 1.0000 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0.5185 \\ 0.7523 \\ 0.3125 \\ 1 \end{bmatrix}. \]
C.4 Other Types of Kriging

If we wish to estimate average (also known as regularised) values over an area or a volume rather than point values, we may use block kriging which can be combined with several other forms of kriging.

If more variables are studied simultaneously the above methods for description of spatial correlation may be extended to handle the spatial covariance between all pairs of variables in the form of cross semi-variograms or cross-covariance functions. Also more variables may be interpolated simultaneously using cokriging. Cokriging is most useful when one variable is sampled on fewer locations than other correlated variables. Universal kriging is a method for the case where the mean value is described by linear combinations of known functions ideally determined by the physics of the problem at hand. Also methods for non-linear kriging exist such as lognormal kriging, multi-Gaussian kriging, rank kriging, indicator kriging and disjunctive kriging. References here are [1], [3], [9], [14].

C.5 Conclusions, kriging

The above sections and examples demonstrate the following properties of kriging:

- Kriging is a type of interpolation that gives us both an estimate based on the spatial structure of the variable in question as expressed by the auto-covariance function (or the semi-variogram) as well as an estimation variance which is minimised.
- The kriging estimator is the best linear unbiased estimator (BLUE) in the sense that it minimises the estimation variance. Also it is exact, i.e., if we interpolate to a point which coincides with an existing sample point, kriging gives the same value as the one measured and the kriging variance is zero.
- The kriging system and the kriging variance depend on the auto-covariance function (or the semi-variogram) and the spatial layout of the sample locations only and not on the actual data values. If an auto-covariance function (or a semi-variogram) is known (or assumed or imposed) this has important potential for minimising the estimation variance in experimental design (i.e., in the planning phase of the spatial lay-out of the sampling scheme).
- The solution of the kriging system implies a statistical distance weighting of the data values which enter into the interpolation. Also for OK, the weights are scaled so they add to one. Furthermore, possible redundancy in the form of clustering of the sample locations is accounted for; the above mentioned screening effect is due to this de-clustering.
- Because of the implicit re-estimation of the mean value for each new point constellation, OK is well suited for situations where the mean is not constant over the study region, i.e., where we don’t have first order stationarity.

Further, the kriging system has a unique solution if and only if the covariance matrix $C$ (Section III-C) is positive definite; this also guarantees a non-negative kriging variance.

The strength of kriging may be attributed to a combination of the above characteristics.

If we choose to formulate the kriging system in terms of the auto-covariance function which is done in this note, we must assume (or impose) second order stationarity, i.e., the same auto-covariance function over the entire study area. The system may also be formulated in terms of the semi-variogram and in this case we must assume the intrinsic hypothesis, i.e., the same semi-variogram over the entire study area.

These assumptions may seem to be a drawback of kriging but if deterministic methods are applied, we implicitly make similar
assumptions. It can hardly be considered as a drawback of geo-
statistical methods that we are forced to consider whether such
assumptions are appropriate.

IV. Final Remarks

A sampling strategy may be based on the dependence of the
kriging variance on the distance to the nearest samples. If the
auto-covariance function (or the semi-variogram) and the sam-
ple locations are known, we can determine the kriging weights
and the kriging variances before the actual sampling takes place.
If the variances become too large in some regions of our study
area we may modify the sample locations to obtain smaller vari-
ances. Also, to obtain a good estimate of the nugget effect which
is an important parameter for the outcome of the kriging process,
it may be an advantage to position some samples close to each
other.

In multivariate studies where the joint behaviour of several
variables is investigated, rather than interpolating the original
variables we may interpolate combinations of them. For in-
stance we may interpolate principal components or factors re-
sulting from a factor analysis or a spatial factor analysis, [19],
[20], [11], [21], [22]. Generel references to multivariate statistics
are for example [23], [24]. [25] is written especially for

Also temporal aspects in connection with the application of
data that vary in both space and time may be important. Spatio-
temporal semi-variograms and spatio-temporal kriging are dealt
with in for example [26], [27]. A GIS for handling of temporal
data is described in [28].

REFERENCES
1979, http://www.stokos.demon.co.uk/.
data sampled on a regional scale,” in Statistics in the Environmental and
Earth Sciences, A. Walden and P. Guttrop, Eds., pp. 283–300. Griffin,
[10] Y. Pannatier, VARIOWIN: Software for Spatial Data Analysis in 2D,
[11] Allan A. Nielsen, Analysis of Regularly and Irregularly Sampled Spa-
tial, Multivariate, and Multi-temporal Data, Ph.D. thesis, Informatics
and Mathematical Modelling, Technical University of Denmark, Lyngby,
Informatics and Mathematical Modelling, Technical University of Den-
South African Workshop on Pattern Recognition, Paul Clilliers, Ed., Si-
mon’s Town, South Africa, 25–26 November 1993, pp. 25–35.
[14] Allan A. Nielsen, “Kriging,” lecture note, Technical University of Den-
[16] M. J. D. Powell, “The theory of radial basis function approximation,” in
Advances in Numerical Analysis II: Wavelets, Subdivision Algorithms and
[18] F. Anton, J. A. Barrentzen, J. Gravesen, and H. Aanaes, Computational Ge-
ometry Processing, lecture note, Technical University of Denmark, 2008.
geochemical data from metavolcanic rocks in the Ben Nevis area, On-
factor analysis of multivariate data,” Computers & Geosciences, vol. 17,
[21] Allan A. Nielsen, Knut Conradson, John L. Pedersen, and Agnete Steen-
felt, “Spatial factor analysis of stream sediment geochemistry data from
South Greenland,” in Proceedings of the Third Annual Conference of the
International Association for Mathematical Geology, Vera Pawlowsky-
[22] Allan A. Nielsen, Knut Conradson, John L. Pedersen, and Agnete Steen-
felt, “Maximum autocorrelation factorial kriging,” in Proceedings of the 6th
International Geostatistics Congress (Geostats 2000), W. J. Kleingeld
and D. G. Krieger, Eds., Cape Town, South Africa, April 2000. Internet
[23] Knut Conradson, Introduktion til Statistik, Informatics and Mathematical
Modelling, Technical University of Denmark, 1984.
ograms with use in agriculture,” in Lecture Notes in Statistics, T. G. Gre-
[27] Annette Kjær Erbsøl and Bjørn Kjær Erbsøl, “On spatio-temporal krig-
ing,” in Proceedings of the Third Annual Conference of the Interna-
tional Association for Mathematical Geology (IAMG’97), Vera Pawlowsky-
[28] Thomas Knudsen, Busstop - A Spatio-Temporal Information System,
Ph.D. thesis, Niels Bohr Institute, Department of Geophysics, University
of Copenhagen, Denmark, 1997.